10.48047/jocaaa.2024.33.08.141

A Deep Dive into Cloud Data Storage Security: Vulnerabilities and Mitigation Techniques

Varun Kumar Singh
Ph.D. Research Scholar
Department of Computer Science
C. S. J. M. University, Kanpur
varunksingh07@gmail.com

Dr. Rashi Agarwal
Faculty of Computer Science
Department of Computer Science
C. S. J. M. University, Kanpur

Dr. Renu Jain
Head of the Department
Department of Computer Science, C. S. J. M. University, Kanpur

***Corresponding Author:** Varun Kumar Singh, Ph.D. Research Scholar**,** Department of Computer Science**,** C. S. J. M. University, Kanpur**,** varunksingh07@gmail.com

Abstract

Cloud computing offers users the ability to store, process, and retrieve data from anywhere and anytime, but it lacks efficient data security. Service providers claim that clients generate and manage security-based personal encryption keys, but this is not always the case. Governments and courts often ask cloud hosting companies to provide user data during crises, and hackers are also trying to gain access to data illegally. This research work proposes a novel approach called the Four Tier Secure Cloud Storage Architecture (FTSCSA) to store and retrieve data economically and securely using multiple clouds. The proposed model focuses on improving data integrity, confidentiality, and availability. The architecture involves four tiers of operations: encryption, encoding, hashing, and distribution to different clouds. The complete file can be retrieved from multiple clouds by performing the four tiers operations in reverse, and any lost or corrupted chunks are regenerated. The proposed FTSCSA model takes less storage space and response time for various file sizes, and performs better for larger files due to its splitting of the file into smaller chunks and storage in different clouds.

Keywords: Sensor, Cloud Computing, Security, Artificial Intelligence, System Architecture.

Introduction

Cloud computing refers to the use of various computing resources provided as a service across a network, storing and accessing data on remote servers rather than relying on local hard drives and private data centers. It allows companies to avoid upfront infrastructure expenses and focus on projects that differentiate their businesses instead. The launch of Salesforce.com in 1999 was one of the early turning points for cloud computing, making it easier for both specialized and general-purpose software companies to deploy programs online [1-3].

Computing is the use of computer hardware and software for various purposes, including designing and building hardware and software systems for various purposes, processing, structuring, and managing information, conducting scientific studies with computers, making computer systems behave intelligently, creating and using communications and entertainment media, and finding and gathering relevant information [4]. There are several types of

10.48047/jocaaa.2024.33.08.141

computing, including traditional computing, personal computing, time-sharing computing, client-server computing, peer-to-peer computing, mobile computing, distributed computing, and virtualization technology. Traditional computing uses physical data centers for different data assets, while personal computing consists of a single machine with all necessary peripherals and complete programs. Time-sharing computing allows multiple users to share a system concurrently, while client-server computing consists of two machines with a client and server machine [5-7]. Peer-to-peer computing offers services using several nodes throughout the network, while mobile computing executes tasks on devices like smartphones and tablets.

Distributed computing involves multiple computers and devices connecting using a common network and communicating over an available network. Distributed computing is a method of computing where different parts of a program are run simultaneously on two or more computers communicating with each other over an available network. In conclusion, cloud computing has evolved significantly due to the development of cloud services, the evolution of computing technologies, and the increasing acceptance of online services by businesses [8-10].

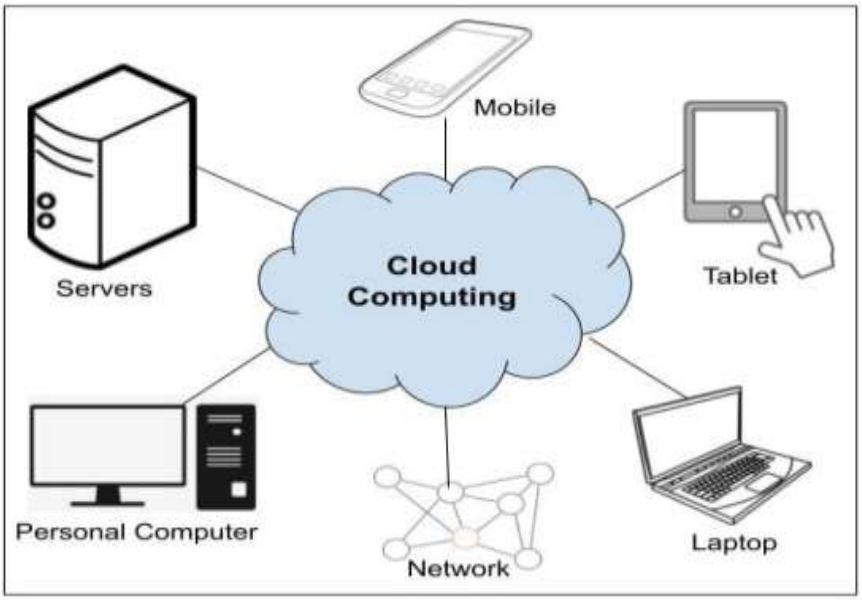

Fig.1 Architecture of Cloud Computing

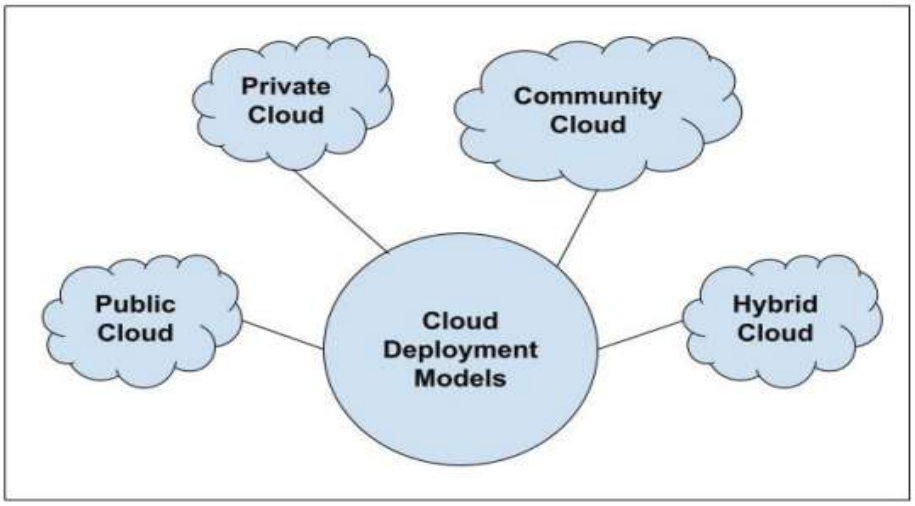

Fig.2 Cloud Deployment Models

10.48047/jocaaa.2024.33.08.141

The CIA (Confidentiality, Integrity & Availability) Model is a crucial aspect of cloud computing security. It ensures user confidence in accessing sensitive data through proper encryption methods. Integrity maintains data integrity by preventing unauthorized modifications or deletions [11-15]. Availability ensures data availability to users through backup and redundancy techniques. The literature review covers encryption technology, erasure code, data partitioning, and multi-cloud technology. The Role-Based Encryption (RBE) technique provides various roles to users and privileges using cross-cloud storage architecture and cryptographic techniques. The RBE system efficiently decrypts data on the client side. A 1024-bit key is generated using DNA-based computing technology to secure large files in the cloud. This method uses user characteristics, MAC addresses, and other personal attributes to generate a secret key. DNA computing, based on the biological concept of DNA, is a cutting-edge emerging field for enhancing data security [16-18].

The SaveMe approach combines public and private cloud storage with client-side security, utilizing encryption and encoding techniques to ensure data security and integrity. It does not require additional server components and offers scalable volume control, data encryption, and redundancy. The DEPSKY system uses encryption, encoding, and replication methods, utilizing four commercial clouds and a protocol to enable users in multiple countries to access the service. This system uses Byzantine quorum system protocols, cryptography, secret sharing, erasure codes, and the diversity of having multiple clouds [19-23].

A safe, compact, reliable, and effective data exchange protocol between media clouds and mobile users is proposed using High Efficiency Video Coding (HEVC) for data concealment. Heroku, a cloud platform, can be used to secure data using AES cryptography. Data security and integrity are checked using a hierarchical identity-based cryptographic approach. Elliptic curve cryptography is used for data encryption in cloud computing due to its minimal key size and energy consumption. Steganography is considered the most effective method for safeguarding communication in the cloud [24-26].

Table-1 Best Cloud Storage Services Providers

Services	Provider	Free Storage	Security
Simple Storage Service (S3)	Amazon	1 Year Free	SSL and automatic encryption
Google Drive	Google	15 GB	128-bit AES, two-step verification
Dropbox	Dropbox, Inc.	2 GB	AES 256-bit encryption, SSL
OneDrive	Microsoft	15 GB	Encryption, two-step verification
Mega	Mega	50 GB	End-to-end Encryption
Tresorit	Tresorit	3 GB	Encryption, TLS
pCloud	pCloud	20 GB	256-bit AES encryption, 4096-bit RSA

The Four Tier Secure Cloud Storage Architecture (FTSCSA) is a specialized system model designed to enhance the security of cloud data. It consists of encryption (AES), encoding (SRC), hashing (MD5), and chunk allocation (CSCA). The AES encryption algorithm

10.48047/jocaaa.2024.33.08.141

prevents hackers in cloud environments, while SRC encoding ensures data availability even if multiple clouds fail. MD5 Hashing maintains data accuracy and consistency. The FTSCSA model offers a quick and effective way to securely store and retrieve data from private or public cloud storages. It combines multiple levels of security to provide an efficient model for cloud users. The model consists of three major modules: front-end module, file storage module, and file retrieval module. The front-end module allows users to add or delete cloud storages, store and retrieve files, and delete files [27]. The model ensures confidentiality and secrecy by using four levels of protection that are impossible for hackers to breach.

The file storage module is the main module in the FTSCSA model, which focuses on storing files securely in multiple clouds. The intercloud or cloud of cloud models use techniques like encryption, encoding, and partitioning processes to ensure data confidentiality and availability. The CIA triad model is the ideal security paradigm, with an extra integrity check using a hashing approach. Encryption is essential for protecting user files from both internal and external attacks. The Advanced Encryption Standard (AES) algorithm is one of the best methods to encrypt files. The first phase of the FTSCSA model uses the symmetric encryption based AES algorithm with a 128-bit key for enhanced security. However, it is difficult to keep and safeguard the key used for the encryption process. The FTSCSA concept uses an automatically generated key to encrypt the file saved in the cloud [28].

In the example given, the phrase "colors are the smiles of nature" is encrypted using the FTSCSA model. The key is expanded using Rijndael's key schedule, generating fresh 128-bit round keys.

This research paper discusses the feasibility of multi-cloud architecture for data storage, focusing on four feasible models: 2X Replication Model, Data Partitioning Model, Cloud-RAID Model, and Regenerating Code Model. The 2X replication model ensures data availability by writing or copying the same data and storing it in different clouds. However, it consumes more storage, cost, time, and bandwidth. Data partitioning ensures data security by dividing data into two big partitions using Maximum Distance Separable (MDS) technique. Each partition is replicated and stored in different clouds. If any partitions are missing during retrieval, they can be regenerated from another replicated partition. The combining process combines both partitions to retrieve the complete file and decrypts to get back the original file [29].

The Cloud-RAID model uses redundant array of independent discs (RAID) technology to store identical data in many locations to safeguard data in case of drive failure. It satisfies Confidentiality and Availability but has a slow writing operation and longer rebuilding times when a cloud fails. The regenerating code model uses a regenerating code technique instead of RAID, which involves encryption of user data, splitting the file into data chunks, creating parity chunks, and recovering data chunks in case of loss. The data chunks and parity chunks are stored across multiple clouds and regenerated from other clouds with the help of parity chunks. In conclusion, the proposed multi-cloud system model satisfies all three principles of the CIA Triad Model, including confidentiality, availability, and integrity [30].

Methodology and Performance Analysis

The FTSCSA model, which complies with the CIA model, offers four tiers of high-end security for cloud files. It ensures all attributes of the CIA model are met while maintaining

10.48047/jocaaa.2024.33.08.141

efficiency and quality. The model's performance is evaluated against other techniques using a cloud computing infrastructure with a HP Proliant DL385 G7 server as a cloud controller. Four well-known public cloud storages, including Google Drive, Microsoft OneDrive, Amazon Drive, and Dropbox, were used to assess the FTSCSA model. The HP Proliant DL385 G7 server is a popular rack-mount server with 48 cores, 32GB RAM, and 8 x 600GB hard disk. It supports standard SFF hot-plug drive bays, SATA Slimline DVD-ROM optical drive, and redundant power supply.

The FTSCSA system is designed for secure cloud storage, allowing only authorized users to access it. Users must authenticate by entering their virtual cloud username and password. The dashboard page displays tasks for secure cloud storage, including Single Cloud Upload & Download for a 2 MB file, Encryption Time, Download, Hash & Decode, Decryption Time, and Total Time taken for Single Cloud vs Multi Cloud upload and download. The FTSCSA architecture allows for multiple cloud storage, with files stored in different partitions. The file is split into blocks x, y, and s and stored in different clouds.

The architecture also allows for storing multiple files, with files available in Google, Amazon, Microsoft Azure, and Rackspace Cloud. The storage space required for different file sizes ranges from 32 MB to 512 MB. The 2x replication and partitioning techniques take up 200% storage space, while CloudRAID consumes 1.66 (166%) times and the FTSCSA model consumes 1.5x (150%) storage space. The total storage required for the FTSCSA model is very low compared to other models, indicating that the FTSCSA model is a more efficient and secure solution for secure cloud storage.

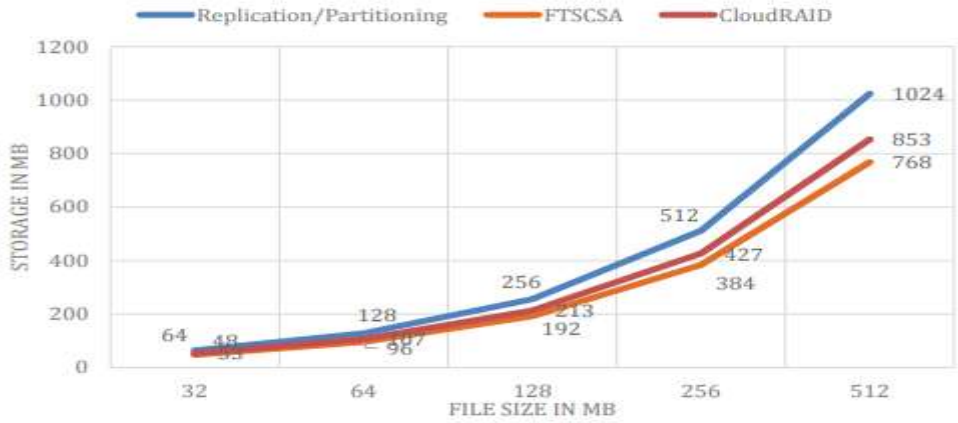

Fig. 3 Storage Requirement for Different Cloud Models

Fig. 4 File Upload Time

10.48047/jocaaa.2024.33.08.141

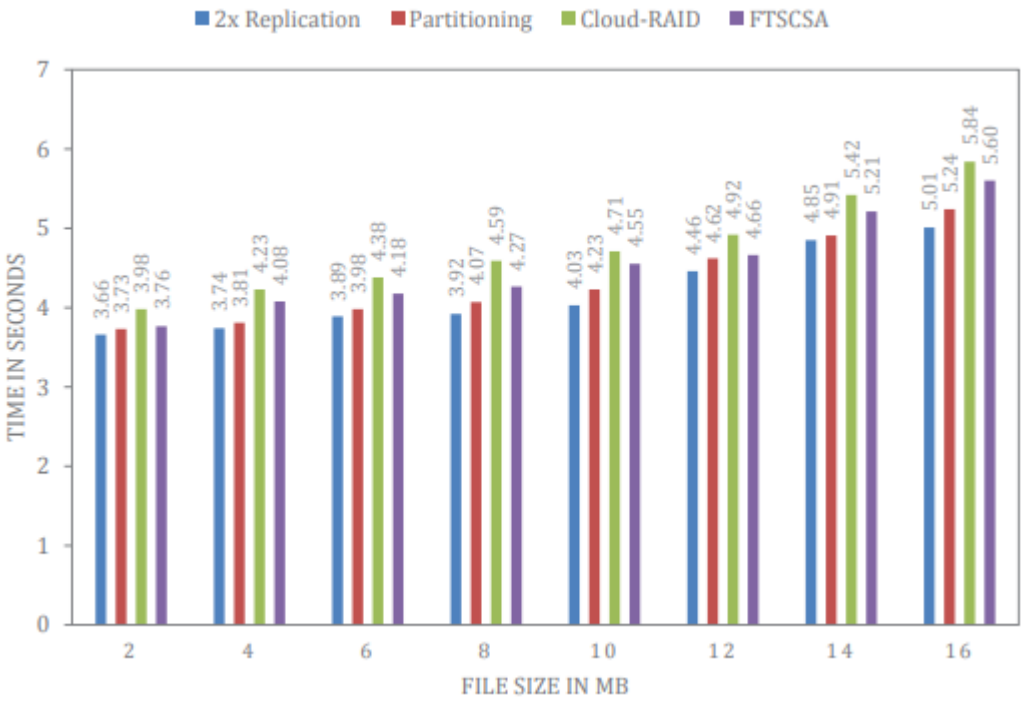

Fig.5 Download Time

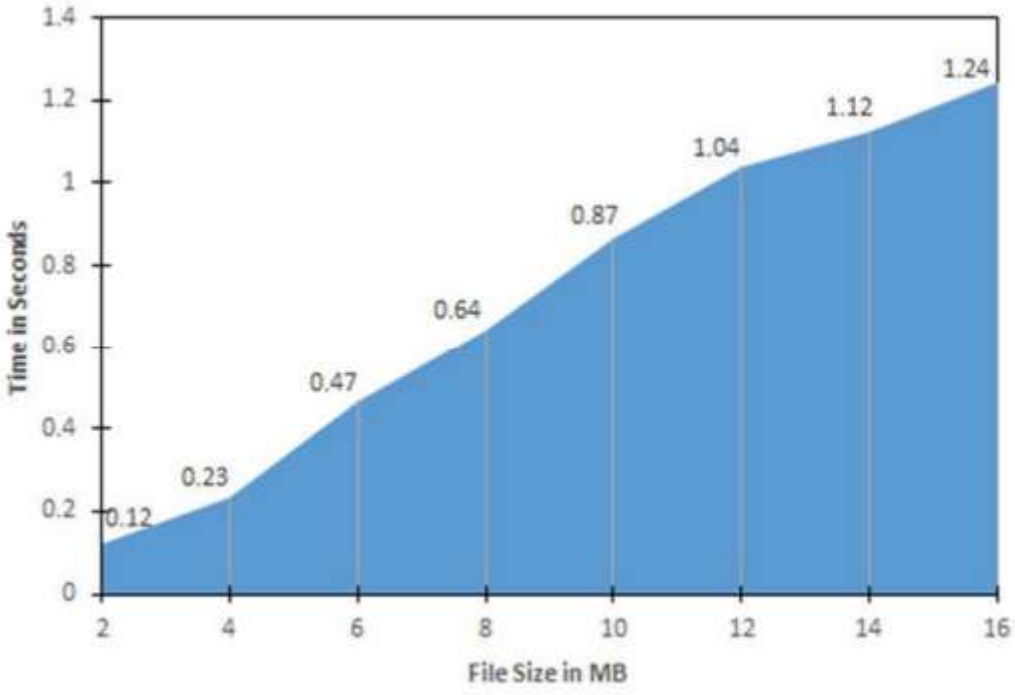

Fig. 6 Rebuilding Time for Single Cloud Failure

10.48047/jocaaa.2024.33.08.141

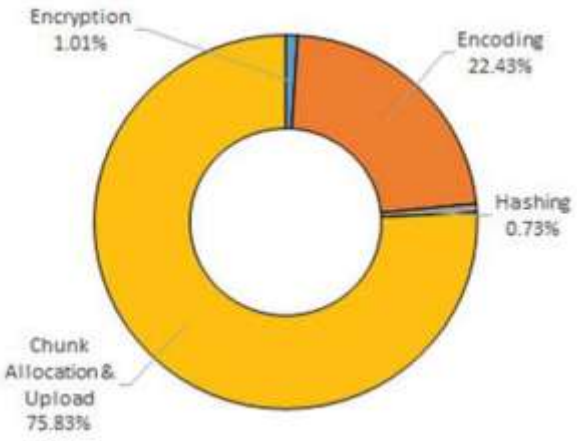

Fig. 7 File Storage time in percentage

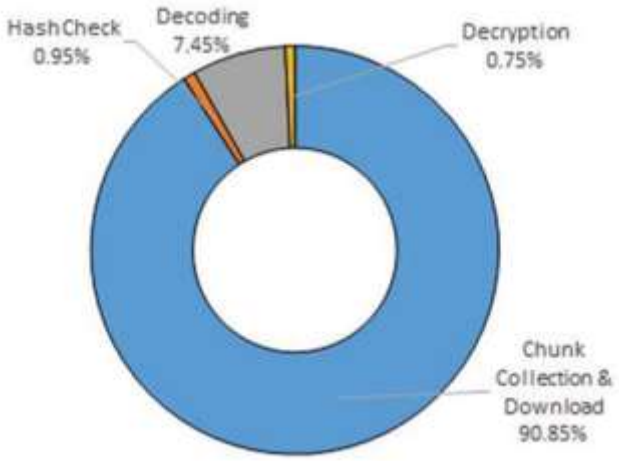

Fig. 8 File Retrieval time in percentage

Conclusion

Cloud computing provides users with the ability to store, process, and retrieve data from anywhere and anytime, but it lacks efficient data security. Service providers claim that clients generate and manage security-based personal encryption keys, but this is not always the case. Governments and courts often ask cloud hosting companies to provide user data during crises, and hackers are also trying to gain access to data illegally. This research work proposes a novel approach called the Four Tier Secure Cloud Storage Architecture (FTSCSA) to store and retrieve data economically and securely using multiple clouds. The proposed model focuses on improving data integrity, confidentiality, and availability. The architecture involves four tiers of operations: encryption, encoding, hashing, and distribution to different clouds. The complete file can be retrieved from multiple clouds by performing the four tiers operations in reverse, and any lost or corrupted chunks are regenerated. The proposed FTSCSA model takes less storage space and response time for various file sizes, and performs better for larger files due to its splitting of the file into smaller chunks and storage

10.48047/jocaaa.2024.33.08.141

in different clouds. The Four Tier Secure Cloud Storage Architecture (FTSCSA) is a system model designed to enhance cloud data security. It consists of encryption (AES), encoding (SRC), hashing (MD5), and chunk allocation (CSCA). The model consists of three major modules: front-end, file storage, and file retrieval. The main module, file storage, stores files securely in multiple clouds. The CIA triad model is the ideal security paradigm, with an extra integrity check using a hashing approach. The model uses the AES algorithm for encryption, but the key is automatically generated. The research paper discusses the feasibility of multi-cloud architecture for data storage, focusing on four feasible models: 2X Replication Model, Data Partitioning Model, Cloud-RAID Model, and Regenerating Code Model.

References

1. Abdullah & Ako Muhamad 2017, 'Advanced encryption standard (AES) algorithm to encrypt and decrypt data', Cryptography and Network Security, vol. 16, pp. 1-11.
2. Adeela Waqar, Asad Raza, Haider Abbas & Muhammad Khurram Khan 2013, 'A framework for preservation of cloud users' data privacy using dynamic reconstruction of metadata', Journal of Network and Computer Applications, vol. 36, pp. 235-248, https://doi.org/10.1016/ j.jnca.2012.09.001 [J
3. Alex, ME & K ishore, R 2017, 'Forensics framework for cloud computing', Computers & Electrical Engineering, vol. 60, pp. 193-205.
4. Alexandros G Dimakis, Brighten Godfrey P, Yunnan Wu, Martin J Wainwright & Kannan Ramchandran 2010, 'Network Coding for Distributed Storage Systems', IEEE Transaction on Information Theory, vol. 56, no. 9, https://doi.org/10.1109/TIT.2010.2054295 [September, 2010].
5. Bai, M, Jiang, S, Zhang, X & Wang, X 2022, 'An efficient skyline query algorithm in the distributed environment', Journal of Computational Science, vol. 58, p. 101524.
6. Balakumar, S & K avitha, AR 2021, 'Quorum-based blockchain network with IPFS to improve data security in IoT network', Studies in Informatics and Control, vol. 30, no. 3, pp. 85-98.
7. Ammar Mohammed Ali & Alaa Kadhim Farhan 2020, 'A Novel Improvement With an Effective Expansion to Enhance the MD5 Hash Function for Verification of a Secure E-Document', IEEE Access, vol. 8, https://doi.org/10.1109/ACCESS.2020.2989050 [May, 2020].
8. Boneh, D, Crescenzo, GD, Ostrovsky, R & Persiano, G 2004, 'Public key encryption with keyword search', In International conference on the theory and applications of cryptographic techniques, Springer, Berlin, Heidelberg, pp. 506-522.
9. Caro, MP, Ali, MS, V ecchio, M & Giaffreda, R 2018, 'Blockchainbased traceability in Agri-Food supply chain management: A practical implementation', In IoT V ertical and Topical Summit on AgricultureTuscany (IOT Tuscany), pp. 1-4.
10. Bessani, Correia M, Quaresma B, Andre F & Sousa P 2013, 'DEP SKY: Dependable and Secure Storage in a Cloud-of-Clouds', ACM Transactions on Storage, vol. 9, no. 4, Article 12, https://doi.org/ 10.1145/2535929 [Nov
11. Chen, R, Mu, Y , Y ang, G, Guo, F & Wang, X 2015, 'Dual-server public-key encryption with keyword search for secure cloud storage', IEEE transactions on information forensics and security, vol. 11, no. 4, pp. 789-798.

10.48047/jocaaa.2024.33.08.141

12. Dong, X , Cao, Z & Shen, J 2019, 'Revocable Public K ey Encryption with Authorized K eyword Search', In IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), pp. 857-860.

13. Galvez, JF, Mejuto, JC & Simal-Gandara, J 2018, 'Future challenges on the use of blockchain for food traceability analysis', TrAC Trends in Analytical Chemistry, vol. 107, pp. 222-232.

14. Geetha, K & K annan, A 2019, 'An efficient information system for providing location based services in network environments', Wireless Personal Communications, vol. 109, no. 4, pp. 2377-2398.

15. Islam, SJ, Chaudhury, ZH & Islam, S 2019, 'A simple and secured cryptography system of cloud computing', In IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), pp. 1-3.

16. K han, A, Y aqoob, A, Sarwar, K , Tahir, M & Ahmed, M 2017, 'Secure logging as a service using reversible watermarking', Procedia Computer Science, vol. 110, pp. 336-343.

17. Bulbul Gupta, Pooja Mittal & Tabish Mufti 2021, 'A Review on Amazon Web Service (AWS), Microsoft Azure & Google Cloud Platform (GCP) Services', Proceedings of the 2nd International Conference on ICT for Digital, Smart, and Sustainable Development, ICIDSSD 2020, Jamia Hamdard, pp. 27-28, DOI: 10.4108/eai.27-2- 2020.2303255.

18. K im, M, Hilton, B, Burks, Z & Reyes, J 2018, 'Integrating blockchain, smart contract-tokens, and IoT to design a food traceability solution', In 2018 IEEE 9th annual information technology, electronics and mobile communication conference (IEMCON), pp. 335-340.

19. K shetri, N 2019, 'Blockchain and the economics of food safety', It Professional, vol. 21, no. 3, pp. 63-66.

20. Lei, K , Du, M, Huang, J & Jin, T 2020, 'Groupchain: Towards a scalable public blockchain in fog computing of IoT services computing', IEEE Transactions on Services Computing, vol. 13, no. 2, pp. 252-262.

21. Chao Yin, Changsheng Xie, Jiguang Wan, Chih-Cheng Hung, Jinjiang Liu & Yihua Lan 2013, 'BMCloud: Minimizing Repair Bandwidth and Maintenance Cost in Cloud Storage', Mathematical Problems in Engineering, vol. 2013, https://doi.org/10.1155/2013/756185 [November 201

22. Liu, Q, Tan, CC, Wu, J & Wang, G 2011, 'Reliable re-encryption in unreliable clouds', In IEEE Global Telecommunications ConferenceGLOBECOM, pp. 1-5.

23. Muthurajkumar, S, Ganapathy, S, V ijayalakshmi, M & K annan, A 2017, 'An intelligent secured and energy efficient routing algorithm for MANETs', Wireless Personal Communications, vol. 96, no. 2, pp. 1753-1769.

24. Osmanoglu, M, Tugrul, B, Dogantuna, T & Bostanci, E 2020, 'An effective yield estimation system based on blockchain technology', IEEE Transactions on Engineering Management, vol. 67, no. 4, pp. 1157-1168.

25. David Sanchez & Montserrat Batet 2017, 'Privacy-preserving data outsourcing in the cloud via semantic data splitting', Computer Communications, https://doi.org/10.1016/j.comcom.2017.06.012 [June, 2017].

10.48047/jocaaa.2024.33.08.141

26. Pichan, A, Lazarescu, M & Soh, ST 2018, 'Towards a practical cloud forensics logging framework', Journal of information security and applications, vol. 42, pp. 18-28.

27. Ray, I, Belyaev, K , Strizhov, M, Mulamba, D & Rajaram, M 2013, 'Secure logging as a service— delegating log management to the cloud', IEEE systems journal, vol. 7, no. 2, pp. 323-334.

28. Selvi, M, Thangaramya, K , Ganapathy, S, K ulothungan, K, K hannah Nehemiah, H & K annan, A 2019, 'An energy aware trust based secure routing algorithm for effective communication in wireless sensor networks', Wireless Personal Communications, vol. 105, no. 4, pp. 1475-1490.

29. Dongmahn Seo, Suhyun Kim & Gyuwon Song 2017, 'Mutual Exclusion Method in Client-Side Aggregation of Cloud Storage', IEEE Transactions on Consumer Electronics, vol. 63, no. 2, pp. 185-190, https://doi.org/10.1109/TCE.2017.014838 [May, 2017].

30. Goda, K & Kitsuregawa, M 2012, 'The History of Storage Systems', Proceedings of the IEEE, vol. 100, no. Special Centennial Issue, pp. 1433-1440, doi: 10.1109/JPROC.2012.2189787 [13 May 2012].